Time-data tradeoff for the sparse and cosparse regularizations of physics-driven inverse problems

Srđan Kitić
Inria, Centre Inria Rennes
Université de Rennes, Rennes, France.

Nancy Bertin
IRISA - CNRS UMR 6074,
Rennes, France.

Rémi Gribonval
Inria, Centre Inria Rennes
Rennes, France.

***Abstract*—We investigate the computational performance of the sparse *vs* cosparse regularizations applied to physics-driven inverse problems, relative to the amount of measurements. Our results show that, despite nominal equivalence of the two models in the given context, the analysis-based optimization benefits from an increase in the volume of available data, while the synthesis one does not.**

I. Introduction

In the context of linear inverse problems, the goal is to estimate a signal $\mathbf{x} \in \mathbb{R}^{\mathrm{n}}$ given measurements $\mathbf{y} = \mathbf{M}\mathbf{x} + \boldsymbol{\varepsilon}$, where $\mathbf{M}$ is a linear observation operator and $\boldsymbol{\varepsilon}$ is additive noise. In most cases, the problem is ill-posed (*e.g.* $\mathbf{M} \in \mathbb{R}^{\mathrm{m}\times\mathrm{n}}$, $\mathrm{m} < \mathrm{n}$) and prior information is needed.

In many problems of interest, the first layer of prior information is driven by physics. Examples include acoustic pressure [1], [2], electric potentials in electroencephalography [3], [4] and nuclear magnetization in magnetic resonance imaging [5]. This type of information can be formulated in two ways. One formulation is through linear partial differential equations. In compact notation:

$$\mathcal{A}(X) = Z, \tag{1}$$

where $\mathcal{A}$ is a differential operator (including the associated boundary conditions) affecting the *continuous* signal X on the entire domain of interest Ω. The right term Z usually represents "sources" or "sinks". Another form of representation is through the superposition principle, or compactly:

$$\mathcal{D}(Z) = X, \tag{2}$$

where $\mathcal{D}$ is an integral operator generated from all solutions $\mathcal{A}(X) = \delta_i, \forall i \in \Omega$. These solutions are known as *Green's functions* [6].

The second layer of information is the sparsity assumption. Indeed, in many cases, the right hand side of (1) will be equal to zero, except at singularity points in Ω. We now consider discretized versions of the above vectors and operators: $X \to \mathbf{x}$, $Z \to \mathbf{z}$, $\mathcal{A} \to \mathbf{A}$ and $\mathcal{D} \to \mathbf{D}$. Accordingly, the vector $\mathbf{z} \in \mathbb{R}^{\mathrm{n}}$ is very sparse.

One way to exploit this prior knowledge is to consider the sparse synthesis [7] data model:

$$\mathbf{z}^* = \arg\min_{\mathbf{z}} F_d(\mathbf{MDz} - \mathbf{y}) + F_r(\mathbf{z}) \text{ where } \mathbf{x}^* = \mathbf{D}\mathbf{z}^*. \tag{3}$$

Here $\mathbf{D}$ represents *the dictionary* of Green's functions, and F_d, F_r are the data-fidelity and (sparsity-promoting) regularizer, respectively.

The alternative solution is obtained by the sparse analysis (*a.k.a. cosparse* [8]) data model:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} F_d(\mathbf{Mx} - \mathbf{y}) + F_r(\mathbf{Ax}), \tag{4}$$

where $\mathbf{A}$ is *the analysis operator*.

For properly chosen boundary conditions, equations (1) and (2) have unique solutions. Consequently, the dictionary and the operator are non-singular with $\mathbf{D} = \mathbf{A}^{-1}$, and the two approaches become nominally equivalent [9]. However, they are **substantially different from a computational point of view**: for most dicretizations, the matrix $\mathbf{A}$ contains $\mathcal{O}(\mathrm{n})$ non-zero elements, whilst the matrix $\mathbf{D}$ is usually fully dense, *i.e.* containing $\mathcal{O}(\mathrm{n}^2)$ non-zeros.

II. Showcase: sound source localization

As an example, consider localizing sound sources given microphone recordings of the acoustic pressure X.

The physical model is given by the acoustic wave equation:

$$\Delta X(\mathbf{r}, t) - \frac{1}{c^2}\frac{\partial^2 X(\mathbf{r}, t)}{\partial t^2} = \sum_{\mathrm{j}} Z(\mathbf{r}, t)\delta(\mathbf{r} - \mathbf{r}_{\mathrm{j}})\delta(t - t_{\mathrm{j}}), \tag{5}$$

where c is the speed of sound, Z represents the contribution of sound sources, while $\mathbf{r}$ and t represent the space and time coordinates.

In the discrete setting, this corresponds to the system of difference equations $\mathbf{Ax} = \mathbf{z}$ (obtained, *e.g.* by the Finite Difference Time Domain - FDTD method [10]). The number of variables is $\mathrm{n} = \mathrm{st}$, where s and t are the numbers of points used to discretize space and time, respectively. The measurements are given by $\mathbf{y} = \mathbf{Mx}$, where the row-reduced identity matrix $\mathbf{M} \in \mathbb{R}^{\mathrm{mt}\times\mathrm{st}}$ models an acquisition system with m microphones randomly distributed in space.

Solving the regularized[1] inverse problems (3) and (4) yields estimates of $\mathbf{x}$ and $\mathbf{z}$. The source locations can then be easily obtained from the support of the source term $\mathbf{z}$.

III. Simulations

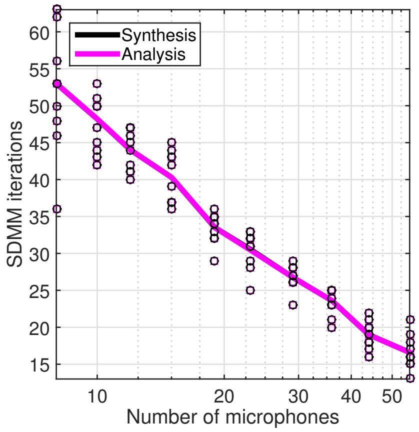
Some recent theoretical studies [12] suggest that the increase of training data enables acceleration of certain machine learning tasks. Inspired by this result, we investigate the behavior of the synthesis and analysis based sound source localization with regards to the amount of measurement data (*i.e.* the number of microphones m).

We simulate the acoustic wave field generated by 5 white noise sources in 2D domain. The wave field is in the form of a rectangular lattice of size $\mathrm{s} = 15 \times 15$, propagating during $\mathrm{t} = 50$ temporal instances. The number of microphones is varied, but such that all sources can be perfectly localized for all values of m.

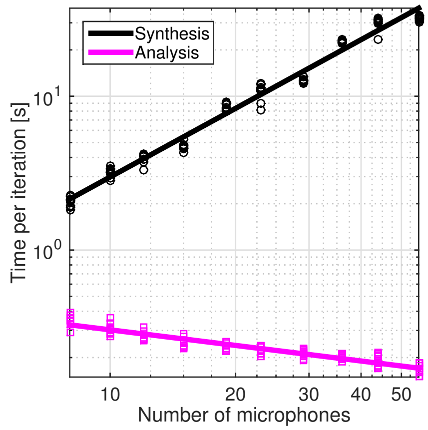
Figure 1(a) confirms that the convergence is indeed accelerated in terms of iteration count for the applied SDMM [13] algorithm. However, the cost per iteration grows for the synthesis-based optimization, as opposed to the analysis-based optimization (figure 1(b)). The overall result given in figure 1(c) is that only the latter benefits from raising the volume of measurement data. In the extreme case, the analysis version is two orders of magnitude faster.

The results are complementary to [14], in the sense that we use a unified algorithmic framework suitable for either synthesis or analysis regularization, and to [15], in the sense that the objective function is not changed. We envision further gains in computational performance by applying a smoothing strategy suggested in the latter article.

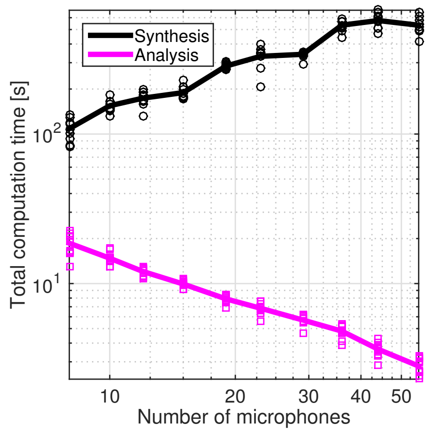
[1] We used the hierarchical $\ell_{2,1}$-norm [11] for F_r and $\iota_{\|\cdot\|_2 \leq \varepsilon}$ for F_d.



(a) Number of iterations *vs* number of microphones.



(b) Processing time per iteration *vs* number of microphones.



(c) Total processing time *vs* number of microphones.

ACKNOWLEDGMENT

This work was supported in part by the European Research Council, PLEASE project (ERC-StG-2011-277906).

REFERENCES

[1] S. Nam and R. Gribonval, "Physics-driven structured cosparse modeling for source localization," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 5397–5400.

[2] I. Dokmanic and M. Vetterli, "Room helps: Acoustic localization with finite elements," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. Ieee, 2012, pp. 2617–2620.

[3] K. Uutela, M. Hämäläinen, and E. Somersalo, "Visualization of magnetoencephalographic data using minimum current estimates," *NeuroImage*, vol. 10, no. 2, pp. 173–180, 1999.

[4] L. Albera, S. Kitic, N. Bertin, G. Puy, and R. Gribonval, "Brain source localization using a physics-driven structured cosparse representation of eeg signals," in *Machine Learning for Signal Processing (MLSP), 2014 IEEE International Workshop on*. IEEE, 2014, pp. 1–6.

[5] M. Davies, G. Puy, V. Pierre, and Y. Wiaux, "A compressed sensing framework for magnetic resonance fingerprinting," *SIAM Journal on Imaging Sciences*, p. 26232656, 2014.

[6] L. C. Evans, "Partial differential equations (graduate studies in mathematics, vol. 19)," *Instructor*, p. 67, 2009.

[7] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM journal on computing*, vol. 24, no. 2, pp. 227–234, 1995.

[8] S. Nam, M. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[9] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse problems*, vol. 23, no. 3, p. 947, 2007.

[10] S. Bilbao, *Numerical Sound Synthesis: Finite Difference Schemes and Simulation in Musical Acoustics*. John Wiley, 2009.

[11] R. Jenatton, J.-Y. Audibert, and F. Bach, "Structured variable selection with sparsity-inducing norms," *The Journal of Machine Learning Research*, vol. 12, pp. 2777–2824, 2011.

[12] S. Shalev-Shwartz and N. Srebro, "SVM optimization: inverse dependence on training set size," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 928–935.

[13] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*. Springer, 2011, pp. 185–212.

[14] R. Giryes, "Sampling in the analysis transform domain," *arXiv preprint arXiv:1410.6558*, 2014.

[15] J. J. Bruer, J. A. Tropp, V. Cevher, and S. Becker, "Time–data tradeoffs by aggressive smoothing," in *Advances in Neural Information Processing Systems*, 2014, pp. 1664–1672.